Indian Journal of Science and Technology, Vol 12(14), DOI: 10.17485/ijst/2019/v12i14/143243, April 2019
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Student Feedback Sentiment Analysis Model using Various Machine Learning Schemes: A Review

Irfan Ali Kandhro[1*], Muhammad Ameen Chhajro[1], Kamlesh Kumar[1], Haque Nawaz Lashari[1] and Usman Khan[2]

[1]Department of Computer Science, Sindh Madressatul Islam University, Karachi, Pakistan
irfan@smiu.edu.pk, ameen.chhajro@smiu.edu.pk, kamlesh@smiu.edu.pk, hnlashari@smiu.edu.pk,
[2]Department of Computer Science, PAF-KIET, Karachi, Pakistan
usman@pafkiet.edu.pk

Abstract

Objectives: The study has been carried out to present the sentiment analysis model for improving the quality of teaching in academic institutions, particularly at universities. The purpose of this study is to explore the different machine learning techniques to identify its importance as well as to raise interest in this research area. In this regard, the student feedback dataset has been collected at the end of the semester Fall-2018 from both Public and Private sector universities of Karachi, through the Google Survey Forms. The dataset contains valuable information about the quality of teaching and learning. **Methods/Statistical Analysis**: The model used the Multinomial Naive Bayes, Stochastic Gradient Decent, Support Vector Machine, Random Forest and Multilayer Perceptron Classifier. The result was analyzed through the evaluation metrics i.e. Confusion Matrix, Precision, Recall and F-score. **Findings**: It is found that the performance of MNB and MLP remained effective as compared to other approaches. It is recommended that MNB and MLP should be used in the research context for the classification of the text. It has great significance for future researchers in sentences and text classification. **Application/Improvements:** The study helps for improving the quality of teaching in education system. And moreover, it will be upgrade by increasing the data samples of neutral comments in dataset.

Keywords: Course Evaluation, Machine Learning, Opinion Mining, Sentiment Analysis, Student Feedback

1. Introduction

The field of Sentiment Analysis emerged a most important field in academic research particularly, in the area of Natural Language Processing (NLP) and Machine Learning during the last decade. Sentiment analysis is focused to examine the problems by identifying the text data, posts, and other contextual information uploaded by users on various social media platforms. The ultimate objective of the sentimental analysis to identify and extract subjective information of various users from the sources and forums regarding their opinions, ideas they have for posts, discussion and conversation that may bring a positive change and lead to help services, businesses, brands and various products and create a room of improvements for them. Which identifies and extracts subjective information in the source material and helping a business to understand the social sentiment of their brand, in this research work the prime focus has been routed towards student's feedback analysis in the form of comments, opinion and reviews regarding the performance of course teacher? Considering the students feedback in this research is because they have diversified learning experiences with various subject teachers and they analyse well in the perspective of teaching skills, methods, knowledge, delivery style and more. Analyzing students' comments feedback using sentiment analysis techniques can categorize the students' positive, negative or neutral sense of feelings regarding the teacher[1].

The most common practice in Educational Institutions is to get feedback from students to analyze their sentiments towards subject teacher in order to improve the performance of the teachers. Basically, this approach of assessment through such feedbacks is usually commenced at the end of the semester with the use of survey forms. However, the subject method does not look efficient and

*Author for correspondence

approach has been found very slow and time-consuming. With the advancement in technology and social media, specifically Facebook, that has widely facilitated in the collection of user feedbacks through various pages and groups. However, extracting the subject information and analysis those feedbacks is a relatively challenging job. This research work provides the best proposed solution and discovers the modern efficient techniques for sentiment analysis applied on student's feedbacks with the usage of machine learning techniques such as Naive Bayes (NB), Stochastic Gradient Descent (SGD), and Support Vector Machines (SVM) respectively. And, evaluate the performances of each subject techniques and perform the comparison analysis with accuracy.

Kaur, proposed methodology based on a survey of classification and opinion mining algorithms. In the researcharea, the sentiment classification is an open field for analysis and experiments. There is a wide range available for subject algorithms. Moreover, Multinomial naive Bayes, SVM, random forest, SGD classifier and multilayer perceptron are dominant approaches for sentiment classification. The suggested model analysis the different domains of text datasets. Like IMDB, Flipkart and Amazon, web and social networks sentiment analysis of words along with context is very much important, so depth study is needed in the required field[2].

The proposed methodology based on emotions of different textual words have been discussed in multidisciplines, like linguistics, psychology and computational linguistics, Currently SA (Sentiment Analysis field has more scope for text classification and categorization. Moreover, the capacity of research work related to emotion words is limited to the perspective of foreign languages learning.

The lack of informative materials, less understanding and complexity of tools hasnarrowlydown the scope of learners, especially in emotion words. The recent survey focuses on the application of opinion mining learning. To achieve the required objectives the author proposed RESOLVE system to develop a context-aware emotion replacement recommendation system, for academic purposes. Using different machine learning techniques, the model can propose synonymous emotion words which are feasible in the context of learners. Significantly, the focus on the used text information data of each emotion words, containing various scenario descriptions, definitions, sentences were given as input in order to develop and help language learner's vocabulary knowledge and words usage. In the education system the proposed model effective and efficient for the writing tasks based on survey questions. The experimental results show that the model achieved a good accuracy on substantial progress in the context of emotion words[3].

The proposed sentiment analysis approach for classifying and identifying the people's thoughts and intentions from the piece of the text. As per researchers views the thoughts can be positive, negative or neutral regarding to the personality. In that sense, the authors do the opinion mining on twitter dataset. The raw data is extracted from various users twitter posts. In order to obtain the results, the machine learning techniques are used for classifying the users' Twitter in the form of text. Moreover, the results show the accuracy of Multinomial Naive Bayes model was state-of-art compared with other ML approaches. The classifier shows that the BJP was a more successful political party in the present time based on public opinion. Model is not restricted to a single domain, but it can be widely used in different domains like business, education, health and the stock market[4].

In proposed a novel method for identifying and classifying the messages of twitter automatically. These texts based short twits are further classified into labels, such as positive and negative with the respect of the query. Also, this is more helpful for customers who are interested to do the survey on opinion mining of different product before purchase, moreover the industries are looking after the all public sentiments of different brands on the web. Therefore, No or limited work has been done on classifying and categorizing the sentiment of microblogging services like twitter messages. The proposed model utilized the machine learning methods for identifying and classifying the short Twitter messages with respect to distant observation. The twitter messages dataset is used in the model for training the emoticons, which are the normally noisy data or labels. And more ever these training samples are richly available to automatically obtain the patterns means.

As suggested author compare the results of machine learning approaches (max entropy, naive bays and support vector machine. the accuracy of these algorithms is greater than 80% while training on emoticons data samples. The main theme of this study is classifying the emoticons data of the twitter messages for distant supervised learning. In this paper, the preprocessing and filtering approach also utilized order to improve the accuracy of the model with respect of emoticons[5].

In proposed presented a new entity-level sentiment approach for analyzing the short twitter messages. Initially, this method adopts a lexicon-based approach to presents entity level opinion mining. The precision of this model is high compared to recall. Additionally, to improve the results of recall require more tweets that are expected to be rigid are discovered automatically by utilizing the data in the result of the lexicon-based approach. Addition to this, text classifier then trained to assign polarities to the entities in the newly identified tweets. Labelled class automatically sets rather than manually, the training examples are given by the lexicon-based approach. The suggested model depicted the Experimental results that significantly expands the recall in a better way and the F-score and outperforms the state-of-the-art baselines[6].

2. Proposed Methodology

M-NB Multinomial Naive Bayes functionally extension of Bayes theorem[7]. Its most popular algorithm for classifications. The performance is good with text and document level opinion mining or sentiment classification. This model is slightly used for data streaming as M-NB classifier directly used for increasing and decreasing the counts. And moreover, these counts are required to estimative the algorithmic and conditional possibilities. The document has used an algorithm for classification that consists of collections of multiple words that represent from each C. P (W|C). Where P shows the possibilities of words in documents that represents the class. The frequency of relative words finds out through the training data as simply estimating the list of the words from documents.

The equation 1 shows estimate prior possibilities as referred P(c) that are required for the classification and other part p(d) is normalization factor that avoids zero frequency issue and Laplace correction for algorithmic possibilities and conditional probability are used replacing the all zero with one.

$$P\left(x_i \mid y\right) = P\left(i|y\right)x_i + \left(1 + P\left(i|y\right)\right)\left(1 - x_i\right) \qquad (1)$$

SGD-Stochastic Gradient Decent is a basic and efficient method for differentiating or differentiable learning in linear classifications under the convex loss functions such as the Logistic regression and SVM. In machine learning the SGD recently get good attention in content scale learning. It's mostly used sparse and large-scale machine problems especially with text classification, moreover, the performance of classifier effective and implementation on text is quite easy. Classifier used Hinge loss or no-differentiable loss with support vector machine, and model adopts the changes over time. Its learning rate is fixed in a simple implementation of the vanilla model. In addition to this, this model loss optimized through L2 penalty regularization approach which is specially used SVM ML techniques. SGD classifier fitted with two arrays: an array X of size [samples, features] holding the training samples, and an array Y of size [samples] holding the target values (class labels) for the training samples: document classifier referred as following formula and optimized used loss function x, and vector represent with b, λ (lambda) is used as regularization parameter. Class labels to (+1,-1). The performance of model compared[8] and moreover not required configuration explicit learning ratio. But there is no improvement is identified latter discussed approaches, another side, and the capability of the algorithm compute with explicit learning rate appeared with twitter data stream task. Many of researchers are used a $\lambda = 0.0001$ and kept the learning rate set for the per-example updates to 0.1 and addition to this, SGD classifier support first-order learning the routine. However, this method iterates the training samples and each cycle the model updates the parameters. Operation of learning parameter also controls the step size in space and similarly intercept is updated but without regularization. Learning rate can be either constant or gradually decaying. For the text classification, the by default learning schedule optimal is given equation (2).

$$w \leftarrow w - \eta\left(\frac{\partial R\left(w\right)}{\partial w} + \frac{\partial L\left(w^T x_i - b, yi\right)}{\partial w}\right) \qquad (2)$$

Support Vector Machine (SVM) is a discriminative classifier as proposed by[9] properly specified by a dividing hyperplane. In training data labels are scattered in two-dimensional space where line is created for dividing the data into two parts. Each part shows the class. Supervised learning model performance is effective on text classification and categorization compare to max-entropy and nave bays theorems. SVM first version introduced by[10] and it used to compute the optimal or best boundaries to separate the negative and positive samples. The performance of this model on training samples are exceptional compared to other machine learning approaches that's suggested authors in[11,12] and some other complexities also discuss in papers as referred[13,14].

And further, to overcome to these complexities require more deep research. The form of the equation defining the decision surface separating the classes is a hyperplane of the form: – w is a weight vector – x is input vector – b is biased equation (3)

$$w \wedge T\, x + b = 0, > 0 = +1 and < 0 = -1 \quad (3)$$

Random Forest or random decision forests[15] as an example approach that main objective of improving and storing classification data in trees form. Ensemble method used for regression, text classification and other tasks that analysis by creating multiple decision trees on training the samples. Model is used correctly for decision trees' habit of overfitting to their training. The arrangement of tree predictors in such way where every single tree patterned values of random vectors are dependent on independently and all trees shared equally across the complete forest. The structure of random forest trees classifier contained a h (x, Ok), k=1 having Ok as individually identically scattered random vectors and every single tree cast a vote for most repeated class or value at input x. In classification (x) is class predication of bth random forest. Where decision of selection of model the perfect model computed with majority vote. As shown in equation (4).

$$\hat{C}_{rf}^{B}(x) = Majority\, Vote\left\{\hat{C}_b(x)\right\}_1^B \quad (4)$$

The Multilayer Perceptron (MLP) is a class of feedforward network. In other words, called nonlinear and robust artificial neural network. In[15] proposedthe MLP consist of three-layer, the input layer, the hidden layer and finally output layer. Every single node or layer contains neurons with nonlinear activation function. And addition to this, MLP used universal activation estimator that contain a single hidden node and having multiple different nonlinear units. And for that reason, learning relation of the variables is good each other. In Multilayer Perceptron (MLP) the flow of data in unidirectional from input to output layer. Its manifold layers and nonlinear activation function for differentiate MLP from a linear perceptron. It can distinguish data that is not linearly separable. Multilayer perceptron that based on Neural Network and it starts with the first input layer contain every single node that having a predictor variable. Input nodes are neurons are interrelated with the neurons in the forward flowing and the next hidden layer with labels. Likewise, the hidden layer neurons relate to the other hidden layer neuron[16-20]. The training MLP classifier by adjusting the weights until it produced correct predication of output as shown equation (5).

$$y_k(x) = \sum_{i=0}^{d} w_{ki} x_i + w_{ko} \quad (5)$$

3. Results and Discussion

In this section, the evaluation of learning models has been performed using dataset. For this purpose, the dataset is built through google forms in which different course evaluation questions are formulated such as quality of teaching, and course contents. However, in order to analyse the student's feedback dataset is split into training and validation sets. And the performance of learning models is assessed through the following metrics:

Accuracy: Accuracy is defined as the ratio between the number of correct predictions made by the model and the number of rows in the dataset. In equation 1 Tp Tn represents true positive and negative whereas Fp and fn show the false positive and negative respectively as shown in equation 6.

$$Accuracy = \frac{Tp + Tn}{Tp + Tn^{+Fp+Fn}} \quad (6)$$

Precision is defined as the ratio between the correct predictions and the total predictions yielded by the system. It can be calculated using the following formula as shown in equation 7.

$$Precision = \frac{Tp}{Tp + Fp} \quad (7)$$

The recall is the ratio between the correct predictions made by the model and the total number of true and false sentiment labels. Recall is computed using below formula as depicted in equation 8.

$$Recall = \frac{Tp}{Tp + Fn} \quad (8)$$

F-Measure is another commonly used metric in the multi-class classification task. It is defined as the geomet-

ric mean of precision and recall. F-measure is an effective metric of measuring the performance of the model as mentioned (9) equation.

$$F-score = \frac{2 * Recall * Acc}{Recall + Acc} \quad (9)$$

Training and testing samples distribution of dataset, Table 1 shows that model training and testing samples with respect labels, there are three labels such as positive, negative and neutral. Samples for positive label 1800 for training and 600 for testing, similarly for negative, the training samples are 900 and 300 for testing, and last, the samples of neutral comments 150 for training and 50 for testing. And moreover Table 2 shows some random samples of the dataset. The confusion matrix indicates for each activity class the distribution of the predictions made by the Sentiment Analysis model Figure 1. The diagonal of the matrix contains the proportion of correctly recognized activity instances, while the off-diagonal elements denote the proportion of activity instances that have been confused with another activity. The off-diagonal elements can be summed up along the recognized activities column to obtain the false positives and along the true. Class row to obtain the false negatives.

The Confusion Matrix of Multinomial Naive Bayes is tested on 1/3 samples and then results are computed with percentage values through positive and negative labels. The Figure 2 of diagonal side represents 87% and

Table 1. Distribution of Student Feedback Dataset in Positive, Negative and Neutral

Separation of the dataset in Training and testing		
Method	Training	Testing
Positive Negative Neutral	1800 900 150	600 300 50

Table 2. Random samples of student feedback dataset with attributes 0, 1 and 2

Dataset Sample with labels 0,1 and 2	
Samples	**Label**
Perfect teacher and best mentor, we like to appreciate you sir, and we never forget the way you taught us & your motivation about the course, moreover, we wish good look, hope you back in next course, THANKYOU SIR	1
Thank you for being a good teacher, you really taught us wonderful concepts about web programming and share the amazing stuff, and moreover, we always remember you with golden words, Good luck	1
The timing of the lab is not good, it's too much lately, for the practical course need A fresh mind, feelings, of course, is boring and sleepy, LOL	0
Kind request sir, give us more practical task & assignments to improve the level of the course to include critical thinking and solutions as it is required in a field of CS research.	2

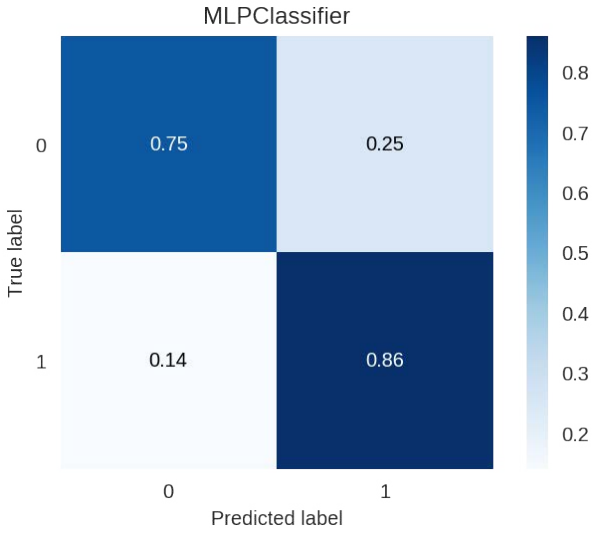

Figure 1. Multilayer perception Confusion Matrix on Student Feedback.

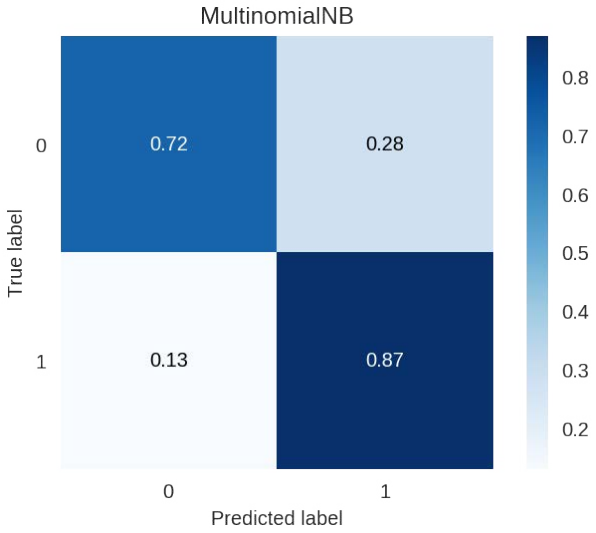

Figure 2. Confusion Matrix of Multinomial Naive Bayes on Student Feedback.

72% true values and on the other side 28% and 13% false values respectively. However, obtained precision and recall in Figure 3 are 87% and 84% respectively. These numerical results provide that model correctly identified the true labels. Similarly, in Figure 4 stochastic gradient descent classifier Model has been tested over 1/3 of comments dataset. However, the accuracy of the model has been evaluated in percentage values. These are obtained through true and false positive, true and false negative outcomes. Therefore, in the confusion matrix, the diagonal side shows 83% and 72% true values and on another side 28% and 17% false values respectively. In Figure 5 precision and recall ratiois 90% and 80% which reveal that it attained high accuracy. Similarly, the same testing dataset has been implemented using support vector classifier. Graphical results in Figure 6 provide that it has been tested over 30% of comments. And confusion matrix represents 83% and 72% true labels and 28% and 17% false labels on the other side. As shown in Figure 7 it achieves a precision of 86% and recall of 82% respectively.

As shown in Figure 8 Model has been tested over 1/3 of comments dataset. However, the accuracy of model has been evaluated in percentage values. These are obtained through true and false positive, true and false negative outcomes. Therefore, in confusion matrix, the diagonal side shows 83% and 48% true values and on another side 52% and 17% false values respectively. As shown in Figure 9 Multilayer perception Model has been tested over 1/3

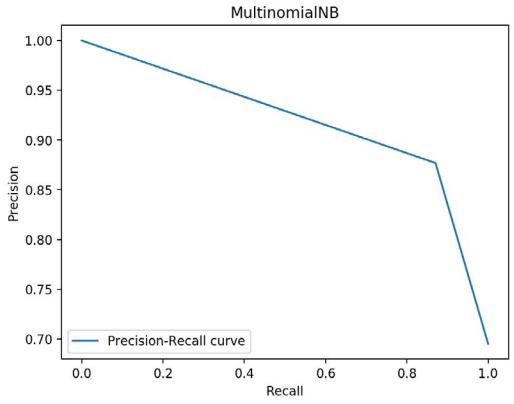

Figure 3. Precision and Recall curve of Multinomial Naive Bayes.

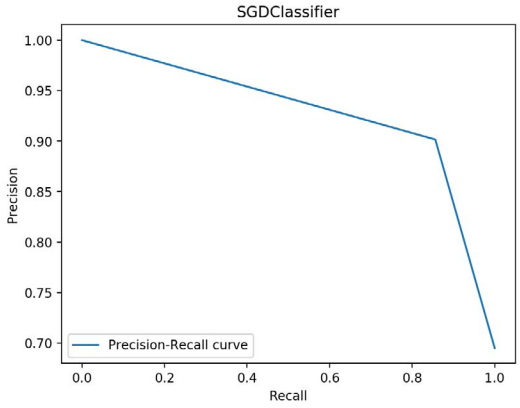

Figure 5. Precision-Recall curve of Stochastic Gradient Descent Classifier.

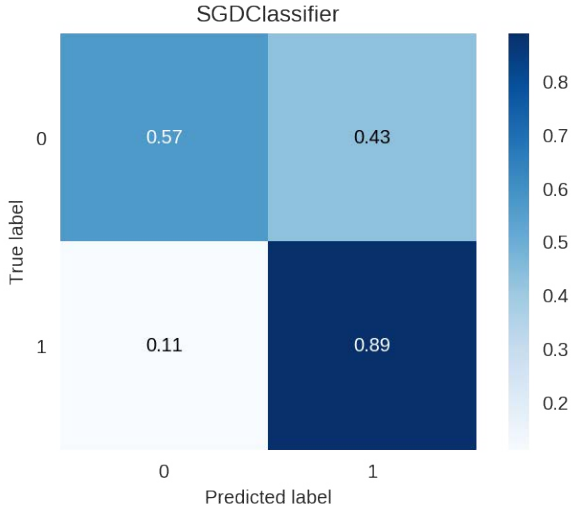

Figure 4. Confusion Matrix of Stochastic Gradient Descent Classifier on Student Feedback.

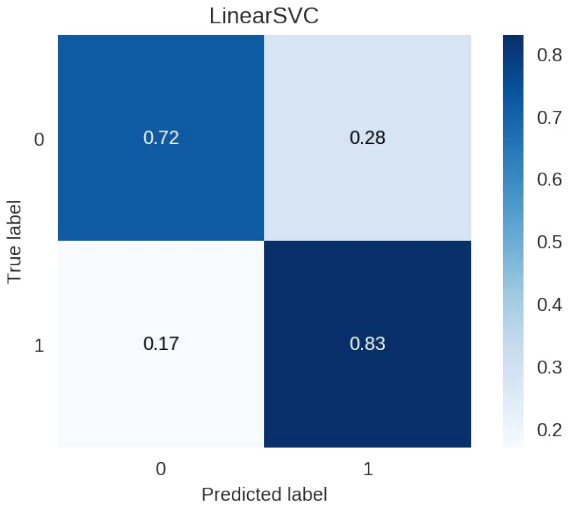

Figure 6. Confusion Matrix of Linear Support Vector Classifier on Student Feedback.

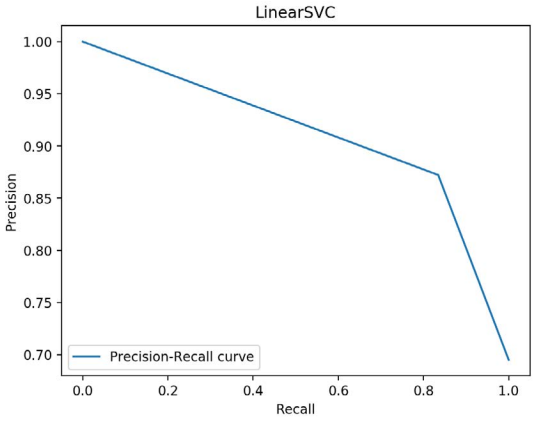

Figure 7. Precision-Recall curve of Linear Support Vector Classifier.

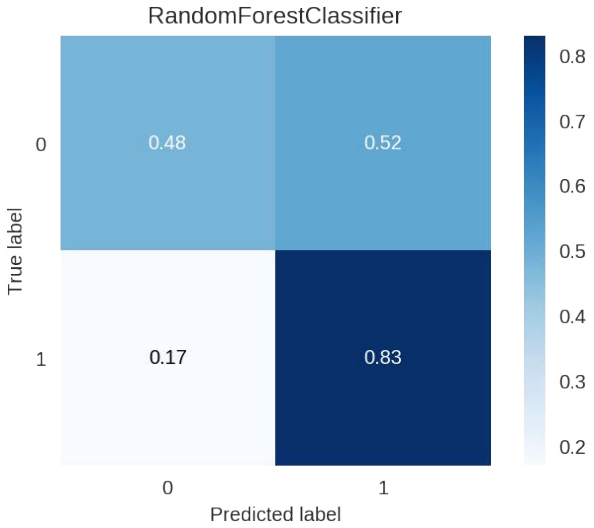

Figure 8. Confusion Matrix of Random Forest Classifier on Student Feedback.

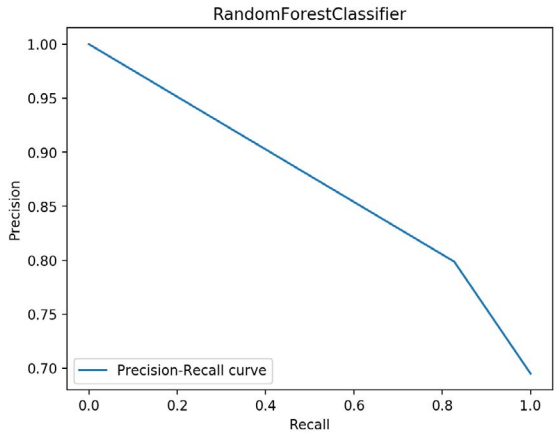

Figure 9. Precision-Recall curve of Random Forest Classifier.

of comments dataset. However, the confusion matrix of multilayer perceptron shows 86% and 75% true labels which is quite a good number as compared to the models. Also, with lesser values of 25% and 14% false labels respectively. Eventually, precision and recall values are 87% and 86% which are quite higher. Figure 10 illustrates that F-score of sentiment analysis method also varies with different tested models. Table 3 shows obtained f-score numerical values are 87% 85%, 85%, 80%, and 87% for MNB, SGD, SVM, Random forest and MLP respectively. The comparative analysis shows that the performance of the MNB and MLP model is quite well over correct sentiment label prediction.

Furthermore, it can be observed from Figures 11 and 12 that accuracy of tested models for sentiment analysis also varies with different machine learning techniques. However, obtained accuracies are 83%, 79%, 80% 72% and 83% for classifier MNB, SGD, SVM Random forest and MLP. From above these figures it can be concluded that the MNB and MLP model performed well in identifying more correct labels as compared to other used methods.

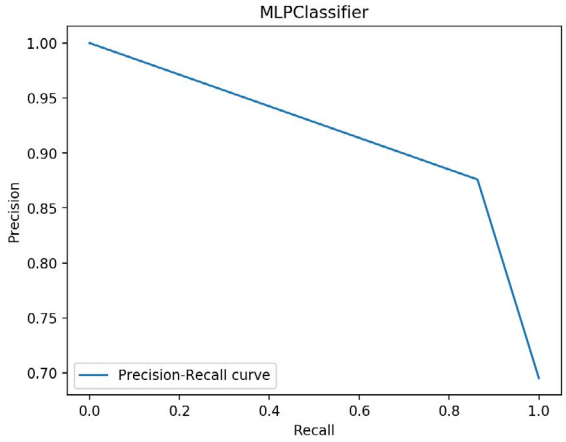

Figure 10. Precision-Recall curve of Multilayer perceptron Classifier.

Table 3. Accuracy and F-score of supervised machine learning methods on Student Feedback

Accuracy Rate and F-score of Machine Learning Methods		
Method	Accuracy	F-score
Multinomial Nave Bays	83%	87%
SGD Classifier	79%	85%
Linear SVC	80%	85%
Random Forest Classifier	72%	80%
MLP Classifier	83%	87%

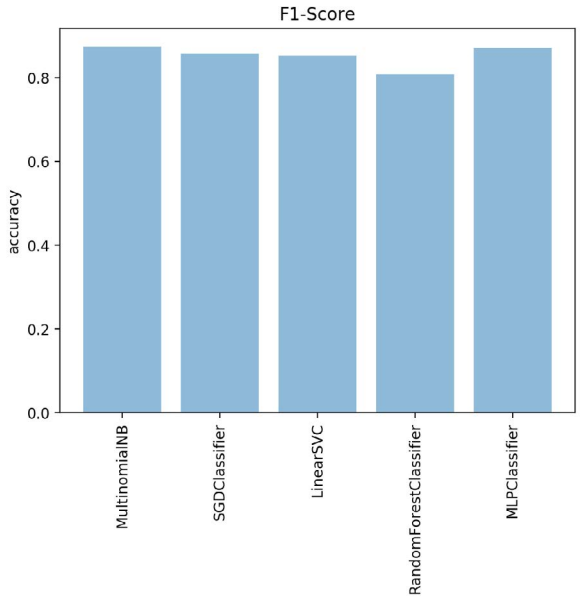

Figure 11. F-Score of sentiment models with various machine learning approaches.

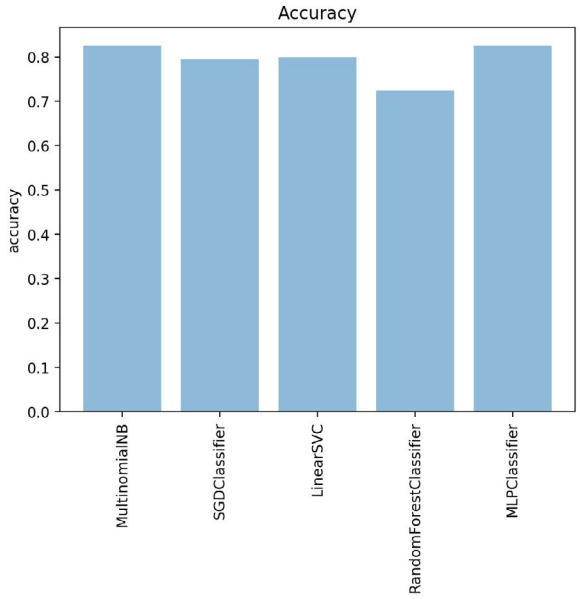

Figure 12. Accuracy of sentiment model with various machine learning approaches.

4. Conclusion

This study has been designed to analyses the discourse commonly used in an academic context. The study used the sentiment analysis survey by incorporating various machine learning techniques. It focuses particularly on students' feedback to strengthen the teaching and learning process at universities. Student feedback dataset is collected at the end of semester fall 2018 from different universities located in Karachi, Sindh. Various machine learning techniques are used to analyses the text data in the SA model. The student feedback dataset split into two parts training and testing; 30% for testing and 70% for training. Moreover, accuracy is evaluated through confusion matrix, Precision, Recall and F-score metrics. The proposed model implemented on Multinomial Naïve Bayes, Stochastic Gradient Decent, Support Vector Machine, Random Forest and Multilayer perceptron Classifier approaches. The results show that the Accuracy of MNB (83.30%), SGD Classifier (83.70%), SVM (80.20%), Random forest (72.20%) and MLP Classifier (83.20%) and similarly F-score of respective techniques are MNB (87%), SGD Classifier (85%), SVM (85%), Random forest (80%) and MLP Classifier (87%). Illustration of graphs and tables show that Multinomial Naive Bayes and MLP Classifier performance is outclassed over all three classifiers namely Random forest, Stochastic Gradient Decent, Support Vector Machine.

5. Future Work

In the future, more preprocessing is required for extracting more features and substrings from dataset. And finally, this work shall be extended in order to implement it over multi lingual.

6. References

1. Stock market sentiment analysis based on machine learning. Available from: https://ieeexplore.ieee.org/document/7877468. Date accessed: 14/10/2016.
2. Sentiment analysis of students' comment using lexicon-based approach. Available from: https://ieeexplore.ieee.org/abstract/document/7959985. Date accessed: 24/05/2017.
3. Ahmad M. Machine Learning Techniques for Sentiment Analysis: A Review. International Journal of Multidisciplinary Space Science and Engineering. 2017; p. 27-32.
4. Using machine learning techniques for sentiment analysis. Available from: https://ddd.uab.cat/pub/tfg/2017/tfg_70824/machine-learning-techniques.pdf. Date accessed: 06/2017.
5. A survey of sentiment analysis techniques. Available from: https://ieeexplore.ieee.org/document/8058315. Date accessed: 10/02/2017.
6. Chen MH, Chen WF, Ku LW. Application of Sentiment Analysis to Language Learning. IEEE Access. 2018; 6:24433-42. https://doi.org/10.1109/ACCESS.2018.2832137

7. Business reviews classification using sentiment analysis. Available from: https://ieeexplore.ieee.org/document/7426090. Date accessed: 21/09/2015.
8. Casting online votes: To predict offline results using sentiment analysis by machine learning classifiers. Available from: https://ieeexplore.ieee.org/document/8203996. Date accessed: 03/07/2017.
9. Twitter sentiment classification using distant supervision. Available from: https://www.researchgate.net/publication/228523135_Twitter_sentiment_classification_using_distant_supervision. Date accessed: 01/2009.
10. Combining lexicon based and learning-based methods for twitter sentiment analysis. Available from: https://www.hpl.hp.com/techreports/2011/HPL-2011-89.html. Date accessed: 2011.
11. Bo P, Lee L. Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval. 2008; 2(1-2):1-135.
12. Cortes C, Vapnik V. Support vector machine. Machine Learning. 1995; 20(3):273-97. https://doi.org/10.1023/A:1022627411411 https://doi.org/10.1007/BF00994018
13. Abbasi A. Affect analysis of web forums and blogs using correlation ensembles. IEEE Transactions on Knowledge and Data Engineering. 2008; 20(9):1168-80. https://doi.org/10.1109/TKDE.2008.51
14. Argamon S. Stylistic text classification using functional lexical features. Journal of the American Society for Information Science and Technology. 2007; 58(6):802-22. https://doi.org/10.1002/asi.20553
15. Kaya GT, Ersoy OK, Kamasak ME. Support vector selection and adaptation for remote sensing classification. IEEE Transactions on Geoscience and Remote Sensing. 2011; 49(6):2071-9. https://doi.org/10.1109/TGRS.2010.2096822
16. Wilson T, Wiebe J, Hwa R. Recognizing strong and weak opinion clauses. Computational intelligence. 2006; 22(2):73-99. https://doi.org/10.1111/j.1467-8640.2006.00275.x
17. Khairnar J, Kinikar M. Machine learning algorithms for opinion mining and sentiment classification. International Journal of Scientific and Research Publications. 2013; 3(6):1-6.
18. Shai SS. Pegasos: Primal estimated sub-gradient solver for svm. Mathematical Programming. 2011; 127(1):3-30. https://doi.org/10.1007/s10107-010-0420-4
19. Leo B. Random forests. Machine Learning. 2001; 45(1):5-32. https://doi.org/10.1023/A:1010933404324
20. Kumar SP, Husain MS. Methodological study of opinion mining and sentiment analysis techniques. International Journal on Soft Computing. 2014; 5(1):11-20. https://doi.org/10.5121/ijsc.2014.5102